

JP01/01127 16.02.01

日本国特許庁

PATENT OFFICE
JAPANESE GOVERNMENT

REC'D 02 MAR 2001

WIPO PCT

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日

Date of Application:

2000年 2月28日

出願番号

Application Number:

特願2000-051466

出願人

Applicant(s):

ソニー株式会社

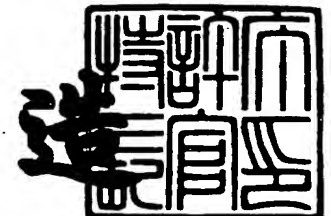
EJU

PRIORITY
DOCUMENTSUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1 (a) OR (b)

2000年12月 8日

特許庁長官
Commissioner,
Patent Office

及川耕造



出証番号 出証特2000-3102528

【書類名】 特許願

【整理番号】 0000008403

【提出日】 平成12年 2月28日

【あて先】 特許庁長官殿

【国際特許分類】 G10L 15/00

【発明者】

 【住所又は居所】 東京都品川区北品川6丁目7番35号 ソニー株式会社
 内

 【氏名】 浅野 康治

【発明者】

 【住所又は居所】 東京都品川区北品川6丁目7番35号 ソニー株式会社
 内

 【氏名】 南野 活樹

【発明者】

 【住所又は居所】 東京都品川区北品川6丁目7番35号 ソニー株式会社
 内

 【氏名】 小川 浩明

【発明者】

 【住所又は居所】 東京都品川区北品川6丁目7番35号 ソニー株式会社
 内

 【氏名】 ヘルムート ルッケ

【特許出願人】

 【識別番号】 000002185

 【氏名又は名称】 ソニー株式会社

 【代表者】 出井 伸之

【代理人】

 【識別番号】 100082131

 【弁理士】

 【氏名又は名称】 稲本 義雄

【電話番号】 03-3369-6479

【手数料の表示】

【予納台帳番号】 032089

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9708842

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 音声認識装置および音声認識方法、並びに記録媒体

【特許請求の範囲】

【請求項 1】 入力された音声に対して、その音声認識結果の尤度を表すスコアを計算し、そのスコアに基づいて、前記音声を認識する音声認識装置であって、

音声認識の対象とする単語群から、既に前記スコアの計算がされた単語に接続する可能性の高い 1 以上の第 1 の単語を選択するとともに、所定の 1 以上の第 2 の単語を選択する選択手段と、

前記選択手段において選択された前記第 1 および第 2 の単語について、前記スコアを計算するスコア計算手段と、

前記スコアに基づいて、前記音声の音声認識結果としての単語列を確定する確定手段と

を備えることを特徴とする音声認識装置。

【請求項 2】 前記選択手段は、音韻数が所定の条件を満たす単語を、前記第 2 の単語として選択する

ことを特徴とする請求項 1 に記載の音声認識装置。

【請求項 3】 前記選択手段は、品詞が所定の条件を満たす単語を、前記第 2 の単語として選択する

ことを特徴とする請求項 1 に記載の音声認識装置。

【請求項 4】 前記音声の特徴量を抽出する抽出手段をさらに備え、

前記選択手段は、前記音声の特徴量を用いて、音声認識の対象とする単語群の各単語について、前記スコアを計算し、そのスコアに基づいて、前記第 1 の単語を選択する

ことを特徴とする請求項 1 に記載の音声認識装置。

【請求項 5】 入力された音声に対して、その音声認識結果の尤度を表すスコアを計算し、そのスコアに基づいて、前記音声を認識する音声認識方法であって、

音声認識の対象とする単語群から、既に前記スコアの計算がされた単語に接続

する可能性の高い 1 以上の第 1 の単語を選択するとともに、所定の 1 以上の第 2 の単語を選択する選択ステップと、

前記選択ステップにおいて選択された前記第 1 および第 2 の単語について、前記スコアを計算するスコア計算ステップと、

前記スコアに基づいて、前記音声の音声認識結果としての単語列を確定する確定ステップと

を備えることを特徴とする音声認識方法。

【請求項 6】 入力された音声に対して、その音声認識結果の尤度を表すスコアを計算し、そのスコアに基づいて、前記音声を認識する音声認識処理を、コンピュータに行わせるプログラムが記録されている記録媒体であって、

音声認識の対象とする単語群から、既に前記スコアの計算がされた単語に接続する可能性の高い 1 以上の第 1 の単語を選択するとともに、所定の 1 以上の第 2 の単語を選択する選択ステップと、

前記選択ステップにおいて選択された前記第 1 および第 2 の単語について、前記スコアを計算するスコア計算ステップと、

前記スコアに基づいて、前記音声の音声認識結果としての単語列を確定する確定ステップと

を備えるプログラムが記録されている

ことを特徴とする記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、音声認識装置および音声認識方法、並びに記録媒体に関し、特に、例えば、音響的な特徴量が不安定な単語を含む音声であっても、少ないリソースで、精度の良い音声認識を行うことができるようにする音声認識装置および音声認識方法、並びに記録媒体に関する。

【0002】

【従来の技術】

図 1 は、従来の音声認識装置の一例の構成を示している。

【 0 0 0 3 】

ユーザが発した音声は、マイク（マイクロフォン）1に入力され、マイク1では、その入力音声、電気信号としての音声信号に変換される。この音声信号は、A D (Analog Digital)変換部2に供給される。A D変換部2では、マイク1からのアナログ信号である音声信号がサンプリング、量子化され、デジタル信号である音声データに変換される。この音声データは、特徴抽出部3に供給される。

【 0 0 0 4 】

特徴抽出部3は、A D変換部2からの音声データについて、適当なフレームごとに音響処理を施し、これにより、例えば、M F C C (Mel Frequency Cepstrum Coefficient)等の特徴量を抽出し、マッチング部4に供給する。なお、特徴抽出部3では、その他、例えば、スペクトルや、線形予測係数、ケプストラム係数、線スペクトル対等の特徴量を抽出することが可能である。

【 0 0 0 5 】

マッチング部4は、特徴抽出部3からの特徴量を用いて、音響モデルデータベース5、辞書データベース6、および文法データベース7を必要に応じて参照しながら、マイク1に入力された音声（入力音声）を、例えば、連続分布HMM法等に基づいて音声認識する。

【 0 0 0 6 】

即ち、音響モデルデータベース5は、音声認識する音声の言語における個々の音素や音節などの音響的な特徴を表す音響モデルを記憶している。ここでは、連続分布HMM法に基づいて音声認識を行うので、音響モデルとしては、例えば、HMM (Hidden Markov Model)が用いられる。辞書データベース6は、認識対象の各単語（語彙）について、その発音に関する情報（音韻情報）が記述された単語辞書を記憶している。文法データベース7は、辞書データベース6の単語辞書に登録されている各単語が、どのように連鎖する（つながる）かを記述した文法規則（言語モデル）を記憶している。ここで、文法規則としては、例えば、文脈自由文法（C F G）や、統計的な単語連鎖確率（N - g r a m）などに基づく規則を用いることができる。

【0007】

マッチング部4は、辞書データベース6の単語辞書を参照することにより、音響モデルデータベース5に記憶されている音響モデルを接続することで、単語の音響モデル（単語モデル）を構成する。さらに、マッチング部4は、幾つかの単語モデルを、文法データベース7に記憶された文法規則を参照することにより接続し、そのようにして接続された単語モデルを用いて、特徴量に基づき、連続分布HMM法によって、マイク1に入力された音声を認識する。即ち、マッチング部4は、特徴抽出部3が出力する時系列の特徴量が観測されるスコア（尤度）が最も高い単語モデルの系列を検出し、その単語モデルの系列に対応する単語列を、音声の認識結果として出力する。

【0008】

即ち、マッチング部4は、接続された単語モデルに対応する単語列について、各特徴量の出現確率を累積し、その累積値をスコアとして、そのスコアを最も高くする単語列を、音声認識結果として出力する。

【0009】

スコア計算は、一般に、音響モデルデータベース5に記憶された音響モデルによって与えられる音響的なスコア（以下、適宜、音響スコアという）と、文法データベース7に記憶された文法規則によって与えられる言語的なスコア（以下、適宜、言語スコアという）とを総合評価することで行われる。

【0010】

即ち、音響スコアは、例えば、HMM法による場合には、単語モデルを構成する音響モデルから、特徴抽出部3が出力する特徴量の系列が観測される確率（出現する確率）に基づいて、単語ごとに計算される。また、言語スコアは、例えば、バイグラムによる場合には、注目している単語と、その単語の直前の単語とが連鎖（接続）する確率に基づいて求められる。そして、各単語についての音響スコアと言語スコアとを総合評価して得られる最終的なスコア（以下、適宜、最終スコアという）に基づいて、音声認識結果が確定される。

【0011】

具体的には、あるN個の単語からなる単語列におけるk番目の単語を w_k とし

て、その単語 w_k の音響スコアを $A(w_k)$ と、言語スコアを $L(w_k)$ と、それぞれ表すとき、その単語列の最終スコア S は、例えば、次式にしたがって計算される。

【0012】

$$S = \Sigma (A(w_k) + C_k \times L(w_k))$$

... (1)

但し、 Σ は、 k を 1 から N に変えてのサメーションをとることを表す。また、 C_k は、単語 w_k の言語スコア $L(w_k)$ にかける重みを表す。

【0013】

マッチング部 4 では、例えば、式 (1) に示す最終スコアを最も大きくする N と、単語列 w_1, w_2, \dots, w_N を求めるマッチング処理が行われ、その単語列 w_1, w_2, \dots, w_N が、音声認識結果として出力される。

【0014】

以上のような処理が行われることにより、図 1 の音声認識装置では、例えば、ユーザが、「ニューヨークに行きたいです」と発話した場合には、「ニューヨーク」、「に」、「行きたい」、「です」といった各単語に、音響スコアおよび言語スコアが与えられ、それらを総合評価して得られる最終スコアが最も大きいときと、単語列「ニューヨーク」、「に」、「行きたい」、「です」が、音声認識結果として出力される。

【0015】

ところで、上述の場合において、辞書データベース 6 の単語辞書に、「ニューヨーク」、「に」、「行きたい」、および「です」の 5 単語が登録されているとすると、これらの 5 単語を用いて構成しうる 5 単語の並びは、 5^5 通り存在する。従って、単純には、マッチング部 4 では、この 5^5 通りの単語列を評価し、その中から、ユーザの発話に最も適合するもの（最終スコアを最も大きくするもの）を決定しなければならない。そして、単語辞書に登録する単語数が増えれば、その単語数分の単語の並びの数は、単語数の単語数乗通りになるから、評価の対象としなければならない単語列は、膨大な数となる。

【0016】

さらに、一般には、発話中に含まれる単語の数は未知であるから、5単語の並びからなる単語列だけでなく、1単語、2単語、・・・からなる単語列も、評価の対象とする必要がある。従って、評価すべき単語列の数は、さらに膨大なものとなるから、そのような膨大な単語列の中から、音声認識結果として最も確からしいものを、計算量および使用するメモリ容量の観点から効率的に決定することは、非常に重要な問題である。

【0017】

計算量およびメモリ容量の効率化を図る方法としては、例えば、音響スコアを求める過程において、その途中で得られる音響スコアが所定の閾値以下となった場合に、そのスコア計算を打ち切るという音響的な枝刈り手法や、言語スコアに基づいて、スコア計算の対象とする単語を絞り込む言語的な枝刈り手法がある。

【0018】

これらの枝刈り手法によれば、スコア計算の対象が、所定の判断基準（例えば、上述したような計算途中の音響スコアや、単語に与えられる言語スコア）に基づいて絞り込まれることで、計算量の削減を図ることができる。しかしながら、その反面、絞り込みを強くすると、即ち、判断基準を厳しくすると、本来、音声認識結果として正しいものまでも枝刈りされてしまい、誤認識が生じることになる。従って、枝刈り手法による場合には、音声認識結果として正しいものが枝刈りされないように、ある程度のマージンをもたせた絞り込みを行う必要があり、このため、計算量を大きく削減することは困難である。

【0019】

また、音響スコアを求める場合に、スコア計算の対象となっているすべての単語について独立に行うと、その計算量が大きくなることから、複数の単語についての音響スコアの計算の一部を共通化（共有化）する方法が提案されている。この共通化の方法としては、単語辞書の単語のうち、その先頭の音韻が同一のものについて、その先頭の音韻から、同一になっている音韻までは、音響モデルを共通に用い、それ以後の異なる音韻には、音響モデルを個々に用いることにより、全体として1つの木構造のネットワークを構成し、これを用いて、音響スコアを求める方法がある。具体的には、例えば、いま、単語「秋田」と「曙」を考え、

「秋田」の音韻情報が「akita」であり、「曙」の音韻情報が「akebono」である
 とすると、単語「秋田」と「曙」の音響スコアは、それぞれの先頭から2番目ま
 での音韻a,kについては兼用で計算される。そして、単語「秋田」の残りの音韻k
 ,i,t,a、および単語「曙」の残りの音韻e,b,o,n,oについては、それぞれ独立に
 音響スコアが計算される。

【0020】

従って、この方法によれば、音響スコアの計算量を大幅に低減することができ
 る。

【0021】

しかしながら、この方法では、共通化されている部分（音響スコアが兼用で計
 算される部分）において、その音響スコアの計算の対象となっている単語を決定
 することができない。即ち、上述の単語「秋田」と「曙」の例でいえば、それぞ
 れの先頭から2番目までの音韻a,kについて音響スコアが計算されている場合は
 、その音響スコアが計算されている単語が、「秋田」であるのか、または「曙」
 であるのかを同定することができない。

【0022】

そして、この場合、「秋田」については、その3番目の音韻iについて音響ス
 コアの計算が開始されたときに、その計算対象が「秋田」であることを同定する
 ことができ、「曙」についても、その3番目の音韻eについての音響スコアの計
 算が開始されたときに、その計算対象が「曙」であることを同定することができ
 る。

【0023】

従って、音響スコアの計算の一部を共通化してしまうと、単語の音響スコアの
 計算の開始時に、その単語を同定することができないため、その単語について、
 言語スコアを考慮することができない。その結果、単語の音響スコアの開始前に
 、上述したような言語的な枝刈り手法を用いることが困難となり、無駄な計算が
 行われることがある。

【0024】

さらに、音響スコアの計算の一部を共通化する場合、単語辞書のすべての単語

を対象として、上述したような木構造のネットワークが構成されるから、これを保持するための大きなメモリ容量が必要となる。

【0025】

また、計算量およびメモリ容量の効率化を図る方法としては、音響スコアを計算する場合に、単語辞書のすべての単語を対象とするのではなく、その音響スコアの計算の対象とする単語を予備的に選択（予備選択）し、その予備選択された単語についてだけ、音響スコアを計算する方法がある。

【0026】

ここで、予備選択の方法は、例えば、L. R. Bahl, S. V. De Gennaro, P. S. Gopalakrishnan and R. L. Mercer, "A Fast Approximate Acoustic Match for Large Vocabulary Speech Recognition", IEEE Trans. Speech and Audio Proc., vol. 1, pp.59-67, 1993等に記載されている。

【0027】

予備選択は、一般に、それほど精度の高くない、簡易的な音響モデルや文法規則を用いて行われる。即ち、予備選択は、単語辞書の単語すべてを対象として行われるため、精度の高い音響モデルや文法規則を用いて予備選択を行うと、リアルタイム性を維持するのに、計算量やメモリ容量といったリソースが多く必要となる。そこで、予備選択は、簡易的な音響モデルや文法規則を用いることで、大語彙を対象とした場合でも、比較的少ないリソースで、高速に行うことが可能となっている。

【0028】

予備選択を行う音声認識装置では、予備選択された単語についてだけマッチング処理を行えば良いので、マッチング処理は、精度の高い音響モデルや文法規則を用いても、少ないリソースで、高速に行うことができる。従って、予備選択を行う音声認識装置は、大語彙を対象として音声認識を行う場合に、特に有用である。

【0029】

【発明が解決しようとする課題】

ところで、予備選択は、ある単語について、特徴量の系列（特徴量系列）を用

いてのマッチング処理が終了し、とりあえず確からしい終点が求められた後に、その終点を始点として、その始点に対応する時刻以後の特徴量系列を用いて行われる。即ち、予備選択は、連続発話された音声に含まれる単語どうしの境界（単語境界）が、最終的に確定していない時点で行われる。

【0030】

従って、予備選択に用いられる特徴量系列の始点や終点が、対応する単語の始点や終点からずれている場合には、その単語の直前の単語や直後の単語の音韻の特徴量を含む特徴量系列や、対応する単語の最初や最後の部分の特徴量が欠けた特徴量系列、即ち、いわば音響的に安定していない特徴量系列を用いて、予備選択が行われることになる。

【0031】

このため、簡易的な音響モデルを用いる予備選択では、発話中に含まれる単語が選択されないことが起こり得る。特に、例えば、日本語の助詞や助動詞、英語の冠詞や前置詞などの音韻数が短い単語については、そのような選択漏れが生じる可能性が高い。

【0032】

そして、予備選択において、正しい単語が選択されない場合には、その単語についてマッチング処理が行われないから、音声認識結果は誤ったものとなる。

【0033】

そこで、予備選択において、単語を選択するときの音響的または言語的な判断基準を緩くして、選択される単語の数を多くする方法や、精度の高い音響モデルおよび文法規則を用いる方法がある。

【0034】

しかしながら、予備選択において、単語を選択するときの音響的または言語的な判断基準を緩くすると、音声認識結果としてそれほど可能性の高くない単語の多くが、マッチング処理の対象となり、予備選択に比較して1単語あたりの負荷の重いマッチング処理に要するリソースが大きく増大する。

【0035】

また、予備選択において、精度の高い音響モデルおよび文法規則を用いる場合

には、予備選択に要するリソースが大きく増大する。

【0036】

本発明は、このような状況に鑑みてなされたものであり、必要なリソースの増加を極力抑えながら、音声認識精度の劣化を防止することができるようにするものである。

【0037】

【課題を解決するための手段】

本発明の音声認識装置は、音声認識の対象とする単語群から、既にスコアの計算がされた単語に接続する可能性の高い1以上の第1の単語を選択するとともに、所定の1以上の第2の単語を選択する選択手段と、選択手段において選択された第1および第2の単語について、スコアを計算するスコア計算手段と、スコアに基づいて、音声の音声認識結果としての単語列を確定する確定手段とを備えることを特徴とする。

【0038】

選択手段には、音韻数が所定の条件を満たす単語を、第2の単語として選択させることができる。

【0039】

また、選択手段には、品詞が所定の条件を満たす単語を、第2の単語として選択させることができる。

【0040】

本発明の音声認識装置には、音声の特徴量を抽出する抽出手段をさらに設けることができ、この場合、選択手段には、音声の特徴量を用いて、音声認識の対象とする単語群の各単語について、スコアを計算させ、そのスコアに基づいて、第1の単語を選択させることができる。

【0041】

本発明の音声認識方法は、音声認識の対象とする単語群から、既にスコアの計算がされた単語に接続する可能性の高い1以上の第1の単語を選択するとともに、所定の1以上の第2の単語を選択する選択ステップと、選択ステップにおいて選択された第1および第2の単語について、スコアを計算するスコア計算ステッ

プと、スコアに基づいて、音声の音声認識結果としての単語列を確定する確定ステップとを備えることを特徴とする。

【0042】

本発明の記録媒体は、音声認識の対象とする単語群から、既にスコアの計算がされた単語に接続する可能性の高い1以上の第1の単語を選択するとともに、所定の1以上の第2の単語を選択する選択ステップと、選択ステップにおいて選択された第1および第2の単語について、スコアを計算するスコア計算ステップと、スコアに基づいて、音声の音声認識結果としての単語列を確定する確定ステップとを備えるプログラムが記録されていることを特徴とする。

【0043】

本発明の音声認識装置および音声認識方法、並びに記録媒体においては、音声認識の対象とする単語群から、既にスコアの計算がされた単語に接続する可能性の高い1以上の第1の単語が選択されるとともに、所定の1以上の第2の単語が選択される。そして、その選択された第1および第2の単語について、スコアが計算され、そのスコアに基づいて、音声の音声認識結果としての単語列が確定される。

【0044】

【発明の実施の形態】

図2は、本発明を適用した音声認識装置の一実施の形態の構成例を示している。なお、図中、図1における場合と対応する部分については、同一の符号を付してあり、以下では、その説明は、適宜省略する。

【0045】

特徴量抽出部3が出力する、ユーザが発した音声の特徴量の系列は、フレーム単位で、制御部11に供給されるようになっており、制御部11は、特徴量抽出部3からの特徴量を、特徴量記憶部12に供給する。

【0046】

また、制御部11は、単語接続情報記憶部16に記憶された単語接続情報を参照し、マッチング部14を制御する。さらに、制御部11は、マッチング部14が、前述した図1のマッチング部4と同様のマッチング処理を行うことにより得

られるマッチング処理結果としての音響スコアや言語スコア等に基づいて、単語接続情報を生成し、その単語接続情報によって、単語接続情報記憶部 1 6 の記憶内容を更新する。また、制御部 1 1 は、単語接続情報記憶部 1 6 に記憶された単語接続情報に基づいて、最終的な音声認識結果を確定して出力する。

【0047】

特徴量記憶部 1 2 は、制御部 1 1 から供給される特徴量の系列を、例えば、ユーザの音声の認識結果が得られるまで記憶する。なお、制御部 1 1 は、音声区間の開始時刻を基準（例えば 0）とする、特徴抽出部 3 が出力する特徴量が得られた時刻（以下、適宜、抽出時刻という）を、その特徴量とともに、特徴量記憶部 1 2 に供給するようになっており、特徴量記憶部 1 2 は、特徴量を、その抽出時刻とともに記憶する。特徴量記憶部 1 2 に記憶された特徴量およびその抽出時刻は、単語予備選択部 1 3 およびマッチング部 1 4 において、必要に応じて参照することができるようになっている。

【0048】

単語予備選択部 1 3 は、マッチング部 1 4 からの要求に応じ、単語接続情報記憶部 1 6、音響モデルデータベース 1 7 A、辞書データベース 1 8 A、および文法データベース 1 9 A を必要に応じて参照しながら、マッチング部 1 4 でマッチング処理の対象とする 1 以上の単語を選択する単語予備選択処理を、特徴量記憶部 1 2 に記憶された特徴量を用いて行う。

【0049】

マッチング部 1 4 は、制御部 1 1 からの制御に基づき、単語接続情報記憶部 1 6、音響モデルデータベース 1 7 B、辞書データベース 1 8 B、および文法データベース 1 9 B を必要に応じて参照しながら、単語予備選択部 1 3 からの単語予備選択処理の結果得られる単語を対象としたマッチング処理を、特徴量記憶部 1 2 に記憶された特徴量を用いて行い、そのマッチング処理の結果を、制御部 1 1 に供給する。

【0050】

単語接続情報記憶部 1 6 は、制御部 1 1 から供給される単語接続情報を、ユーザの音声の認識結果が得られるまで記憶する。

【0051】

ここで、単語接続情報は、最終的な音声認識結果の候補となる単語列を構成する単語どうしの接続（連鎖または接続）関係を表すもので、各単語の音響スコアおよび言語スコア、並びに各単語に対応する発話の開始時刻および終了時刻も含んでいる。

【0052】

即ち、図3は、単語接続情報記憶部16に記憶される単語接続情報を、グラフ構造を用いて示している。

【0053】

図3の実施の形態において、単語接続情報としてのグラフ構造は、単語を表すアーク（図3において、○印どうしを結ぶ線分で示す部分）と、単語どうしの境界を表すノード（図3において○印で示す部分）とから構成されている。

【0054】

ノードは、時刻情報を有しており、この時刻情報は、そのノードに対応する特徴量の抽出時刻を表す。上述したように、抽出時刻は、音声区間の開始時刻を0とする、特徴抽出部3が出力する特徴量が得られた時刻であるから、図3において、音声区間の開始、即ち、最初の単語の先頭に対応するノードNode₁が有する時刻情報は0となる。ノードは、アークの始端および終端となるが、始端のノード（始端ノード）、または終端のノード（終端ノード）が有する時刻情報は、それぞれ、そのノードに対応する単語の発話の開始時刻、または終了時刻となる。

【0055】

なお、図3では、左から右方向が、時間の経過を表しており、従って、あるアークの左右にあるノードのうち、左側のノードが始端ノードとなり、右側のノードが終端ノードとなる。

【0056】

アークは、そのアークに対応する単語の音響スコアおよび言語スコアを有しており、このアークが、終端ノードとなっているノードを始端ノードとして、順次接続されていくことにより、音声認識結果の候補となる単語の系列が構成されていく。

【0057】

即ち、制御部11においては、まず最初に、音声区間の開始を表すノードNode₁に対して、音声認識結果として確からしい単語に対応するアークが接続される。図3の実施の形態では、「今日」に対応するアークArc₁、「いい」に対応するアークArc₆、および「天気」に対応するArc₁₁が接続されている。なお、音声認識結果として確からしい単語かどうかは、マッチング部14において求められる音響スコアおよび言語スコアに基づいて決定される。

【0058】

そして、以下、同様にして、「今日」に対応するアークArc₁の終端である終端ノードNode₂、「いい」に対応するアークArc₆の終端である終端ノードNode₇、「天気」に対応するArc₁₁の終端である終端ノードNode₁₂それぞれに対して、同様に、確からしい単語に対応するアークが接続されていく。

【0059】

以上のようにしてアークが接続されていくことで、音声区間の開始を始点として、左から右方向に、アークとノードで構成される1以上のパスが構成されて行くが、例えば、そのパスのすべてが、音声区間の最後（図3の実施の形態では、時刻T）に到達すると、制御部11において、音声区間の開始から最後までに形成された各パスについて、そのパスを構成するアークが有している音響スコアおよび言語スコアが累積され、最終スコアが求められる。そして、例えば、その最終スコアが最も高いパスを構成するアークに対応する単語列が、音声認識結果として確定されて出力される。

【0060】

具体的には、例えば、図3において、ノードNode₁から、「今日」に対応するアークArc₁、ノードNode₂、「は」に対応するアークArc₂、ノードNode₃、「いい」に対応するアークArc₃、ノードNode₄、「天気」に対応するアークArc₄、ノードNode₅、「ですね」に対応するアークArc₅、およびノードNode₆で構成されるパスについて、最も高い最終スコアが得られた場合には、単語列「今日」、「は」、「いい」、「天気」、「ですね」が、音声認識結果として出力されることになる。

【0061】

なお、上述の場合には、音声区間内にあるノードについて、必ずアークを接続して、音声区間の開始から最後にまで延びるパスを構成するようにしたが、このようなパスを構成する過程において、それまでに構成されたパスについてのスコアから、音声認識結果として不適当であることが明らかであるパスに関しては、その時点で、パスの構成を打ち切る（その後に、アークを接続しない）ようにすることが可能である。

【0062】

また、上述のようなパスの構成ルールに従えば、1つのアークの終端が、次に接続される1以上のアークの始端ノードなり、基本的には、枝葉が広がるように、パスが構成されて行くが、例外的に、1つのアークの終端が、他のアークの終端に一致する場合、つまり、あるアークの終端ノードと、他のアークの終端ノードとが同一のノードに共通化される場合がある。

【0063】

即ち、文法規則としてバイグラムを用いた場合には、別のノードから延びる2つのアークが、同一の単語に対応するものであり、さらに、その単語の発話の終了時刻も同一であるときには、その2つのアークの終端は一致する。

【0064】

図3において、ノードNode₇を始端として延びるアークArc₇、およびノードNode₁₃を始端として延びるアークArc₁₃は、いずれも「天気」に対応するものであり、その発話の終了時刻も同一であるため、その終端ノードは、同一のノードNode₈に共通化されている。

【0065】

なお、ノードの共通化は行わないようにすることも可能であるが、メモリ容量の効率化の観点からは、行うのが好ましい。

【0066】

また、図3では、文法規則としてバイグラムを用いているが、その他、例えば、トライグラム等を用いる場合も、ノードの共通化は可能である。

【0067】

さらに、単語接続情報記憶部 16 に記憶されている単語接続情報は、単語予備選択部 13 およびマッチング部 14 において、必要に応じて参照することができるようになっている。

【0068】

図 2 に戻り、音響モデルデータベース 17A および 17B は、基本的には、図 1 の音響モデルデータベース 5 において説明したような音響モデルを記憶している。

【0069】

但し、音響モデルデータベース 17B は、音響モデルデータベース 17A よりも精度の高い処理が可能な高精度の音響モデルを記憶している。即ち、音響モデルデータベース 17A において、各音素や音節について、例えば、前後のコンテキストに依存しない 1 パターンの音響モデルだけが記憶されているとすると、音響モデルデータベース 17B には、各音素や音節について、例えば、前後のコンテキストに依存しない音響モデルの他、単語間にまたがるコンテキストに依存する音響モデル、つまり、クロスワードモデルや、単語内のコンテキストに依存する音響モデルも記憶されている。

【0070】

辞書データベース 18A および 18B は、基本的には、図 1 の辞書データベース 6 において説明したような単語辞書を記憶している。

【0071】

即ち、辞書データベース 18A および 18B の単語辞書には、同一セットの単語が登録されている。但し、辞書データベース 18B の単語辞書は、辞書データベース 18A の単語辞書よりも精度の高い処理が可能な高精度の音韻情報を記憶している。即ち、辞書データベース 18A の単語辞書には、例えば、各単語に対して、1 通りの音韻情報（読み）だけ登録されているとすると、辞書データベース 18B の単語辞書には、例えば、各単語に対して、複数通りの音韻情報が登録されている。

【0072】

具体的には、例えば、単語「お早う」に対して、辞書データベース 18A の単

語辞書には、1通りの音韻情報「おはよう」だけが、辞書データベース18Bの単語辞書には、「おはよう」の他、「おはよー」や「おはよ」が、それぞれ音韻情報として登録されている。

【0073】

文法データベース19Aおよび19Bは、基本的には、図1の文法データベース7において説明したような文法規則を記憶している。

【0074】

但し、文法データベース19Bは、文法データベース19Aよりも精度の高い処理が可能な高精度の文法規則を記憶している。即ち、文法データベース19Aが、例えば、ユニグラム（単語の生起確率）に基づく文法規則を記憶しているとすると、文法データベース19Bは、例えば、バイグラム（直前の単語との関係を考慮した単語の生起確率）や、トライグラム（直前の単語およびそのさらに1つ前の単語との関係を考慮した単語の生起確率）、文脈自由文法等に基づく文法規則を記憶している。

【0075】

以上のように、音響モデルデータベース17Aには、各音素や音節について、1パターンの音響モデルが、音響モデルデータベース17Bには、各音素や音節について、複数パターンの音響モデルが、それぞれ記憶されている。また、辞書データベース18Aには、各単語について、1通りの音韻情報が、辞書データベース18Bには、各単語について、複数通りの音韻情報が、それぞれ記憶されている。そして、文法データベース19Aには、簡易な文法規則が、文法データベース19Bには、精度の高い文法規則が、それぞれ記憶されている。

【0076】

これにより、音響モデルデータベース17A、辞書データベース18A、および文法データベース19Aを参照する単語予備選択部13では、それほど精度は高くないが、多くの単語を対象として、迅速に、音響スコアおよび言語スコアを求めることができるようになっている。また、音響モデルデータベース17B、辞書データベース18B、および文法データベース19Bを参照するマッチング部14では、ある程度の数の単語を対象として、迅速に、精度の高い音響スコア

および言語スコアを求めることができるようになっている。

【0077】

なお、ここでは、音響モデルデータベース17Aおよび17Bそれぞれに記憶させる音響モデルの精度について優劣を設けるようにしたが、音響モデルデータベース17Aおよび17Bには、いずれにも、同一の音響モデルを記憶させることができ、この場合、音響モデルデータベース17Aおよび17Bは、1つの音響モデルデータベースに共通化することができる。同様に、辞書データベース18Aおよび18Bの単語辞書それぞれの記憶内容や、文法データベース19Aおよび19Bそれぞれの文法規則も、同一にすることができる。

【0078】

次に、図4のフローチャートを参照して、図2の音声認識装置による音声認識処理について説明する。

【0079】

ユーザが発話を行うと、その発話としての音声は、マイク1およびAD変換部2を介することにより、デジタルの音声データとされ、特徴抽出部3に供給される。特徴抽出部3は、そこに供給される音声データから、音声の特徴量を、フレームごとに順次抽出し、制御部11に供給する。

【0080】

制御部11は、何らかの手法で音声区間を認識するようになっており、音声区間においては、特徴抽出部3から供給される特徴量の系列を、各特徴量の抽出時刻と対応付けて、特徴量記憶部12に供給して記憶させる。

【0081】

さらに、制御部11は、音声区間の開始後、ステップS1において、音声区間の開始を表すノード（以下、適宜、初期ノードという）を生成し、単語接続情報記憶部16に供給して記憶させる。即ち、制御部11は、ステップS1において、図3におけるノードNode₁を、単語接続情報記憶部16に記憶させる。

【0082】

そして、ステップS2に進み、制御部11は、単語接続情報記憶部16の単語接続情報を参照することで、途中ノードが存在するかどうかを判定する。

【 0 0 8 3 】

即ち、上述したように、図 3 に示した単語接続情報においては、終端ノードに、アークが接続されていくことにより、音声区間の開始から最後にまで延びるパスが形成されて行くが、ステップ S 2 では、終端ノードのうち、まだアークが接続されておらず、かつ、音声区間の最後にまで到達していないものが、途中ノード（例えば、図 3 におけるノード Node₈ や、Node₁₀、Node₁₁）として検索され、そのような途中ノードが存在するかどうか判定される。

【 0 0 8 4 】

なお、上述したように、音声区間は何らかの手法で認識され、さらに、終端ノードに対応する時刻は、その終端ノードが有する時刻情報を参照することで認識することができるから、アークが接続されていない終端ノードが、音声区間の最後に到達していない途中ノードであるかどうかは、音声区間の最後の時刻と、終端ノードが有する時刻情報とを比較することで判定することができる。

【 0 0 8 5 】

ステップ S 2 において、途中ノードが存在すると判定された場合、ステップ S 3 に進み、制御部 1 1 は、情報接続情報の中に存在する途中ノードのうちの 1 つを、それに接続するアークとしての単語を決定するノード（以下、適宜、注目ノードという）として選択する。

【 0 0 8 6 】

即ち、制御部 1 1 は、情報接続情報の中に 1 つの途中ノードしか存在しない場合には、その途中ノードを、注目ノードとして選択する。また、制御部 1 1 は、情報接続情報の中に複数の途中ノードが存在する場合には、その複数の途中ノードのうちの 1 つを注目ノードとして選択する。具体的には、制御部 1 1 は、例えば、複数の途中ノードそれぞれが有する時刻情報を参照し、その時刻情報が表す時刻が最も古いもの（音声区間の開始側のもの）、または最も新しいもの（音声区間の終わり側のもの）を、注目ノードとして選択する。あるいは、また、制御部 1 1 は、例えば、初期ノードから、複数の途中ノードそれぞれに至るまでのパスを構成するアークが有する音響スコアおよび言語スコアを累積し、その累積値（以下、適宜、部分累積スコアという）が最も大きくなるパス、または小さくな

るパスの終端になっている途中ノードを、注目ノードとして選択する。

【0087】

その後、制御部11は、注目ノードが有する時刻情報を開始時刻としてマッチング処理を行う旨の指令（以下、適宜、マッチング処理指令という）を、マッチング部14に出力する。

【0088】

マッチング部14は、制御部11からマッチング処理指令を受信すると、注目ノード、およびそれが有する時刻情報を、単語予備選択部13に供給し、単語予備選択処理を要求して、ステップS4に進む。

【0089】

ステップS4では、単語予備選択部13は、マッチング部14から、単語予備選択処理の要求を受信すると、注目ノードに接続されるアークとなる単語の候補を選択する単語予備選択処理を、辞書データベース18Aの単語辞書に登録された単語を対象として行う。

【0090】

即ち、単語予備選択部13は、言語スコアおよび音響スコアを計算するのに用いる特徴量の系列の開始時刻を、注目ノードが有する時刻情報から認識し、その開始時刻以降の、必要な特徴量の系列を特徴量記憶部12から読み出す。さらに、単語予備選択部13は、辞書データベース18Aの単語辞書に登録された各単語の単語モデルを、音響モデルデータベース17Aの音響モデルを接続することで構成し、その単語モデルに基づき、特徴量記憶部12から読み出した特徴量の系列を用いて、音響スコアを計算する。

【0091】

また、単語予備選択部13は、各単語モデルに対応する単語の言語スコアを、文法データベース19Aに記憶された文法規則に基づいて計算する。即ち、単語予備選択部13は、各単語の言語スコアを、例えばユニグラムに基づいて求める。

【0092】

なお、単語予備選択部13においては、単語接続情報を参照することにより、

各単語の音響スコアの計算を、その単語の直前の単語（注目ノードが終端となっているアークに対応する単語）に依存するクロスワードモデルを用いて行うことが可能である。

【 0 0 9 3 】

また、単語予備選択部 1 3 においては、単語接続情報を参照することにより、各単語の言語スコアの計算を、その単語が、その直前の単語と連鎖する確率を規定するバイグラムに基づいて行うことが可能である。

【 0 0 9 4 】

単語予備選択部 1 3 は、以上のようにして、各単語について音響スコアおよび言語スコアを求めると、その音響スコアおよび言語スコアを総合評価したスコアを、以下、適宜、単語スコアという）を求め、その上位 L 個を、マッチング処理の対象とする単語として、マッチング部 1 4 に供給する。

【 0 0 9 5 】

さらに、単語予備選択部 1 3 は、ステップ S 4 において、辞書データベース 1 8 A に登録されている所定の 1 以上の単語を、マッチング処理の対象とする単語として選択し、マッチング部 1 4 に供給する。

【 0 0 9 6 】

即ち、単語予備選択部 1 3 は、辞書データベース 1 8 A に登録されている単語のうち、音素数または音韻数が、所定値以下の短い単語、および所定の品詞の単語（例えば、英語における前置詞や冠詞、日本語における助詞や助動詞など）などの、一般に発話時間が短い単語（以下、適宜、特定単語という）を、その音響スコアや言語スコアに関係なく選択し、マッチング処理の対象とする単語として、マッチング部 1 4 に供給する。従って、本実施の形態では、特定単語は、必ず、マッチング処理の対象とされる。

【 0 0 9 7 】

マッチング部 1 4 は、単語予備選択部 1 3 から、単語スコアに基づいて選択された L 個の単語と、単語スコアに関係ない所定の条件に基づいて選択された特定単語を受信すると、ステップ S 5 において、それらの単語を対象として、マッチング処理を行う。

【 0 0 9 8 】

即ち、マッチング部 1 4 は、言語スコアおよび音響スコアを計算するのに用いる特徴量の系列の開始時刻を、注目ノードが有する時刻情報から認識し、その開始時刻以降の、必要な特徴量の系列を特徴量記憶部 1 2 から読み出す。さらに、マッチング部 1 4 は、辞書データベース 1 8 B を参照することで、単語予備選択部 1 3 からの単語の音韻情報を認識し、その音韻情報に対応する音響モデルを、音響モデルデータベース 1 7 B から読み出して接続することで、単語モデルを構成する。

【 0 0 9 9 】

そして、マッチング部 1 4 は、上述のようにして構成した単語モデルに基づき、特徴量記憶部 1 2 から読み出した特徴量系列を用いて、単語予備選択部 1 3 からの単語の音響スコアを計算する。なお、マッチング部 1 4 においては、単語接続情報を参照することにより、単語の音響スコアの計算を、クロスワードモデルに基づいて行うようにすることが可能である。

【 0 1 0 0 】

さらに、マッチング部 1 4 は、文法データベース 1 9 B を参照することで、単語予備選択部 1 3 からの単語の言語スコアを計算する。即ち、マッチング部 1 4 は、例えば、単語接続情報を参照することにより、単語予備選択部 1 3 からの単語の直前の単語と、さらにその前の単語を認識し、トライグラムに基づく確率から、単語予備選択部 1 3 からの単語の言語スコアを求める。

【 0 1 0 1 】

マッチング部 1 4 は、以上のようにして、単語予備選択部 1 3 からの L 個の単語と、特定単語のすべて（以下、適宜、これらをまとめて、選択単語という）について、その音響スコアおよび言語スコアを求め、ステップ S 6 に進む。ステップ S 6 では、選択単語それぞれについて、その音響スコアおよび言語スコアを総合評価した単語スコアが求められ、その単語スコアに基づいて、単語接続情報記憶部 1 6 に記憶された単語接続情報が更新される。

【 0 1 0 2 】

即ち、ステップ S 6 では、マッチング部 1 4 は、選択単語について単語スコア

を求め、例えば、その単語スコアを所定の閾値と比較すること等によって、注目ノードに接続するアークとしての単語を、選択単語の中から絞り込む。そして、マッチング部 1 4 は、その絞り込みの結果残った単語を、その音響スコア、言語スコア、およびその単語の終了時刻とともに、制御部 1 1 に供給する。

【0 1 0 3】

なお、単語の終了時刻は、音響スコアを計算するのに用いた特徴量の抽出時刻から認識される。また、ある単語について、その終了時刻としての蓋然性の高い抽出時刻が複数得られた場合には、その単語については、各終了時刻と、対応する音響スコアおよび言語スコアとのセットが、制御部 1 1 に供給される。

【0 1 0 4】

制御部 1 1 は、上述のようにしてマッチング部 1 4 から供給される単語の音響スコア、言語スコア、および終了時刻を受信すると、マッチング部 1 4 からの各単語について、単語接続情報記憶部 1 6 に記憶された単語接続情報（図 3）における注目ノードを始端ノードとして、アークを延ばし、そのアークを、終了時刻の位置に対応する終端ノードに接続する。さらに、制御部 1 1 は、各アークに対して、対応する単語、並びにその音響スコアおよび言語スコアを付与するとともに、各アークの終端ノードに対して、対応する終了時刻を時刻情報として与える。そして、ステップ S 2 に戻り、以下、同様の処理が繰り返される。

【0 1 0 5】

以上のように、単語接続情報は、マッチング部 1 4 の処理結果に基づいて、逐次更新されるので、単語予備選択部 1 3 およびマッチング部 1 4 は、常時、単語接続情報を利用して処理を行うことが可能となる。

【0 1 0 6】

なお、制御部 1 1 は、単語接続情報を更新する際に、可能であれば、上述したような終端ノードの共通化を行う。

【0 1 0 7】

一方、ステップ S 2 において、途中ノードが存在しないと判定された場合、ステップ S 7 に進み、制御部 1 1 は、単語接続情報を参照することで、その単語接続情報として構成された各パスについて、単語スコアを累積することで、最終ス

コアを求め、例えば、その最終スコアが最も大きいパスを構成するアークに対応する単語列を、ユーザの発話に対する音声認識結果として出力して、処理を終了する。

【0108】

以上のように、単語予備選択部13において、音響スコアおよび言語スコアに基づいて、音声認識結果として確からしい単語を選択する他、例えば、日本語の助詞や助動詞、英語の冠詞や前置詞、その他の音韻数が短い、音響的な特徴量が不安定な単語も選択し、マッチング部14において、それらの単語をマッチング処理の対象とするようにしたので、音響的な特徴量が不安定な単語が、単語予備選択部13で選択されないことによる、音声認識精度の劣化を防止することができる。

【0109】

さらに、この場合、単語予備選択部13において、単語を選択するときの音響的または言語的な判断基準を緩くしたり、精度の高い音響モデルおよび文法規則を用いているわけではないので、処理に必要なリソースの増加を極力低減することができる。

【0110】

また、音響的な特徴量が不安定な、音韻数の短い単語が、必ずマッチング処理の対象とされるため、単語予備選択部13において、音響スコアや言語スコアに基づいて選択される単語は、音響的な特徴量が比較的安定している、音韻数の長い単語だけとなる。従って、単語予備選択部13では、より簡易な音響モデルや文法規則を用いても、正しい単語の選択漏れが生じないこととなり、その結果、単語予備選択部13の処理に必要なリソースを低減しながら、音声認識精度を向上させることができる。

【0111】

さらに、単語予備選択部13において、音響スコアや言語スコアに基づいて選択される単語は、音響的な特徴量が比較的安定している、音韻数の長い単語だけとなることから、単語を選択するときの音響的または言語的な判断基準として、より厳しいものを用い、音響スコアや言語スコアに基づいて選択される単語の数

を少なくとも、正しい単語の選択漏れが生じないこととなり、その結果、マッチング部 1 4 の処理に必要なリソースを低減しながら、音声認識精度を向上させることができる。

【 0 1 1 2 】

次に、上述した一連の処理は、ハードウェアにより行うこともできるし、ソフトウェアにより行うこともできる。一連の処理をソフトウェアによって行う場合には、そのソフトウェアを構成するプログラムが、汎用のコンピュータ等にインストールされる。

【 0 1 1 3 】

そこで、図 5 は、上述した一連の処理を実行するプログラムがインストールされるコンピュータの一実施の形態の構成例を示している。

【 0 1 1 4 】

プログラムは、コンピュータに内蔵されている記録媒体としてのハードディスク 1 0 5 や ROM 1 0 3 に予め記録しておくことができる。

【 0 1 1 5 】

あるいはまた、プログラムは、フロッピーディスク、CD-ROM(Compact Disc Read Only Memory)、MO(Magneto optical)ディスク、DVD(Digital Versatile Disc)、磁気ディスク、半導体メモリなどのリムーバブル記録媒体 1 1 1 に、一時的あるいは永続的に格納（記録）しておくことができる。このようなりムーバブル記録媒体 1 1 1 は、いわゆるパッケージソフトウェアとして提供することができる。

【 0 1 1 6 】

なお、プログラムは、上述したようなりムーバブル記録媒体 1 1 1 からコンピュータにインストールする他、ダウンロードサイトから、デジタル衛星放送用の人工衛星を介して、コンピュータに無線で転送したり、LAN(Local Area Network)、インターネットといったネットワークを介して、コンピュータに有線で転送し、コンピュータでは、そのようにして転送されてくるプログラムを、通信部 1 0 8 で受信し、内蔵するハードディスク 1 0 5 にインストールすることができる。

【0117】

コンピュータは、CPU(Central Processing Unit)102を内蔵している。CPU102には、バス101を介して、入出力インタフェース110が接続されており、CPU102は、入出力インタフェース110を介して、ユーザによって、キーボードや、マウス、マイク等で構成される入力部107が操作等されることにより指令が入力されると、それにしたがって、ROM(Read Only Memory)103に格納されているプログラムを実行する。あるいは、また、CPU102は、ハードディスク105に格納されているプログラム、衛星若しくはネットワークから転送され、通信部108で受信されてハードディスク105にインストールされたプログラム、またはドライブ109に装着されたリムーバブル記録媒体111から読み出されてハードディスク105にインストールされたプログラムを、RAM(Random Access Memory)104にロードして実行する。これにより、CPU102は、上述したフローチャートにしたがった処理、あるいは上述したブロック図の構成により行われる処理を行う。そして、CPU102は、その処理結果を、必要に応じて、例えば、入出力インタフェース110を介して、LCD(Liquid Crystal Display)やスピーカ等で構成される出力部106から出力、あるいは、通信部108から送信、さらには、ハードディスク105に記録等させる。

【0118】

ここで、本明細書において、コンピュータに各種の処理を行わせるためのプログラムを記述する処理ステップは、必ずしもフローチャートとして記載された順序に沿って時系列に処理する必要はなく、並列的あるいは個別に実行される処理（例えば、並列処理あるいはオブジェクトによる処理）も含むものである。

【0119】

また、プログラムは、1のコンピュータにより処理されるものであっても良いし、複数のコンピュータによって分散処理されるものであっても良い。さらに、プログラムは、遠方のコンピュータに転送されて実行されるものであっても良い。

【0120】

なお、マッチング部14でスコア計算の対象となる単語は、単語予備選択部1

3においてあらかじめ選択されているから、マッチング部14による各単語のスコア計算は、前述したような、音響スコアの計算の一部を共通化する木構造のネットワークを構成せずに、各単語ごとに独立して行うことができる。この場合、マッチング部14が各単語についてスコア計算を行うために確保するメモリ容量を小さく抑えることができる。さらに、この場合、単語のスコア計算を開始するときに、その単語が、どの単語であるのかを同定することができるから、前述したような、単語を同定することができないことによって無駄な計算が行われることを防止することができる。

【0121】

また、マッチング部14によるスコア計算は、各単語ごとに、時間的に独立して行うことができ、この場合、スコア計算に要するメモリ容量を使い回すことにより、必要とするメモリ容量を小さく抑えることができる。

【0122】

なお、図2に示した音声認識装置は、例えば、音声によってデータベースの検索を行う場合や、各種の機器の操作を行う場合、各機器へのデータ入力を行う場合、音声対話システム等に適用可能である。より具体的には、例えば、音声による地名の問合せに対して、対応する地図情報を表示するデータベース検索装置や、音声による命令に対して、荷物の仕分けを行う産業用ロボット、キーボードの代わりに音声入力によりテキスト作成を行うディクテーションシステム、ユーザとの会話を行うロボットにおける対話システム等に適用可能である。

【0123】

また、本実施の形態では、単語予備選択部13において、音韻数や品詞に基づいて、特定単語とする単語を選択するようにしたが、その他、例えば、特定単語とする単語は、あらかじめ定めておいて、他の単語とは区別して、単語辞書に登録しておいても良い。

【0124】

さらに、本実施の形態では、単語予備選択部13において、音響スコアおよび言語スコアに基づいて、L個の単語を選択するようにしたが、L個の単語は、その他、例えば、音響スコアだけや、言語スコアだけに基づいて選択することも可

能である。

【0125】

【発明の効果】

本発明の音声認識装置および音声認識方法、並びに記録媒体によれば、音声認識の対象とする単語群から、既にスコアの計算がされた単語に接続する可能性の高い1以上の第1の単語が選択されるとともに、所定の1以上の第2の単語が選択される。そして、その選択された第1および第2の単語について、スコアが計算され、そのスコアに基づいて、音声の音声認識結果としての単語列が確定される。従って、第2の単語が選択されないことによる、音声認識精度を劣化を防止することができる。

【図面の簡単な説明】

【図1】

従来の音声認識装置の一例の構成を示すブロック図である。

【図2】

本発明を適用した音声認識装置の一実施の形態の構成例を示すブロック図である。

【図3】

単語接続情報を説明するための図である。

【図4】

図2の音声認識装置の処理を説明するためのフローチャートである。

【図5】

本発明を適用したコンピュータの一実施の形態の構成例を示すブロック図である。

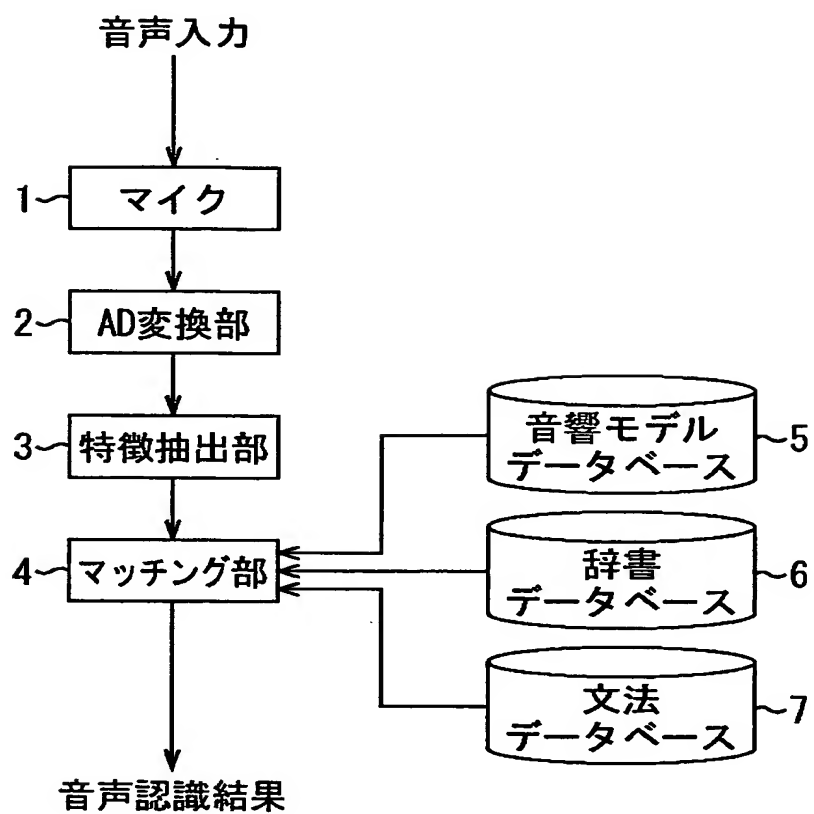
【符号の説明】

1 マイク, 2 AD変換部, 3 特徴抽出部, 11 制御部, 12 特徴量記憶部, 13 単語予備選択部, 14 マッチング部, 16 単語接続情報記憶部, 17A, 17B 音響モデルデータベース, 18A, 18B 辞書データベース, 19A, 19B 文法データベース, 101 バス, 102 CPU, 103 ROM, 104 RAM, 105 ハードディスク

ク, 1 0 6 出力部, 1 0 7 入力部, 1 0 8 通信部, 1 0 9 ドラ
イブ, 1 1 0 入出力インタフェース, 1 1 1 リムーバブル記録媒体

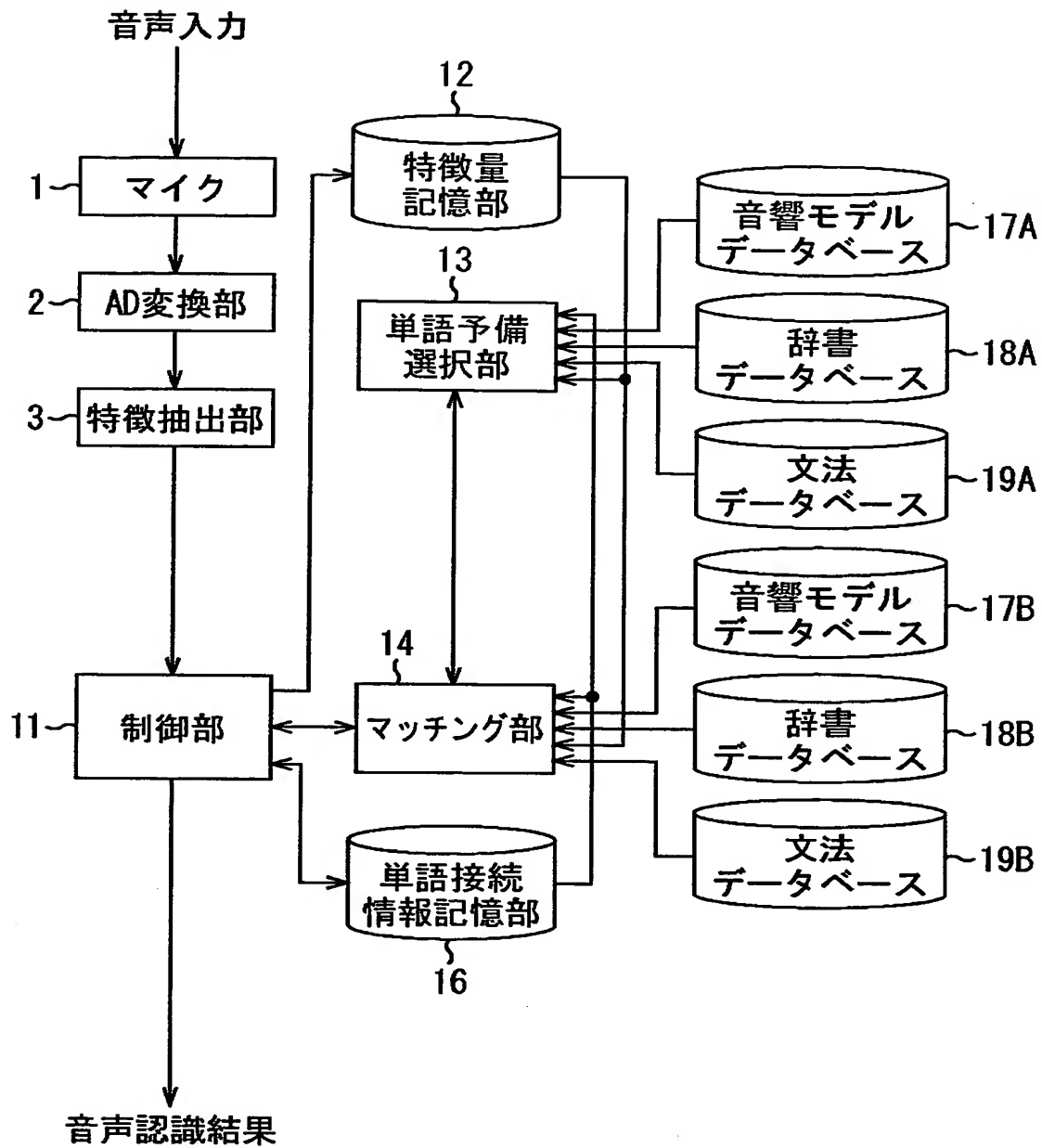
【書類名】図面

【図1】

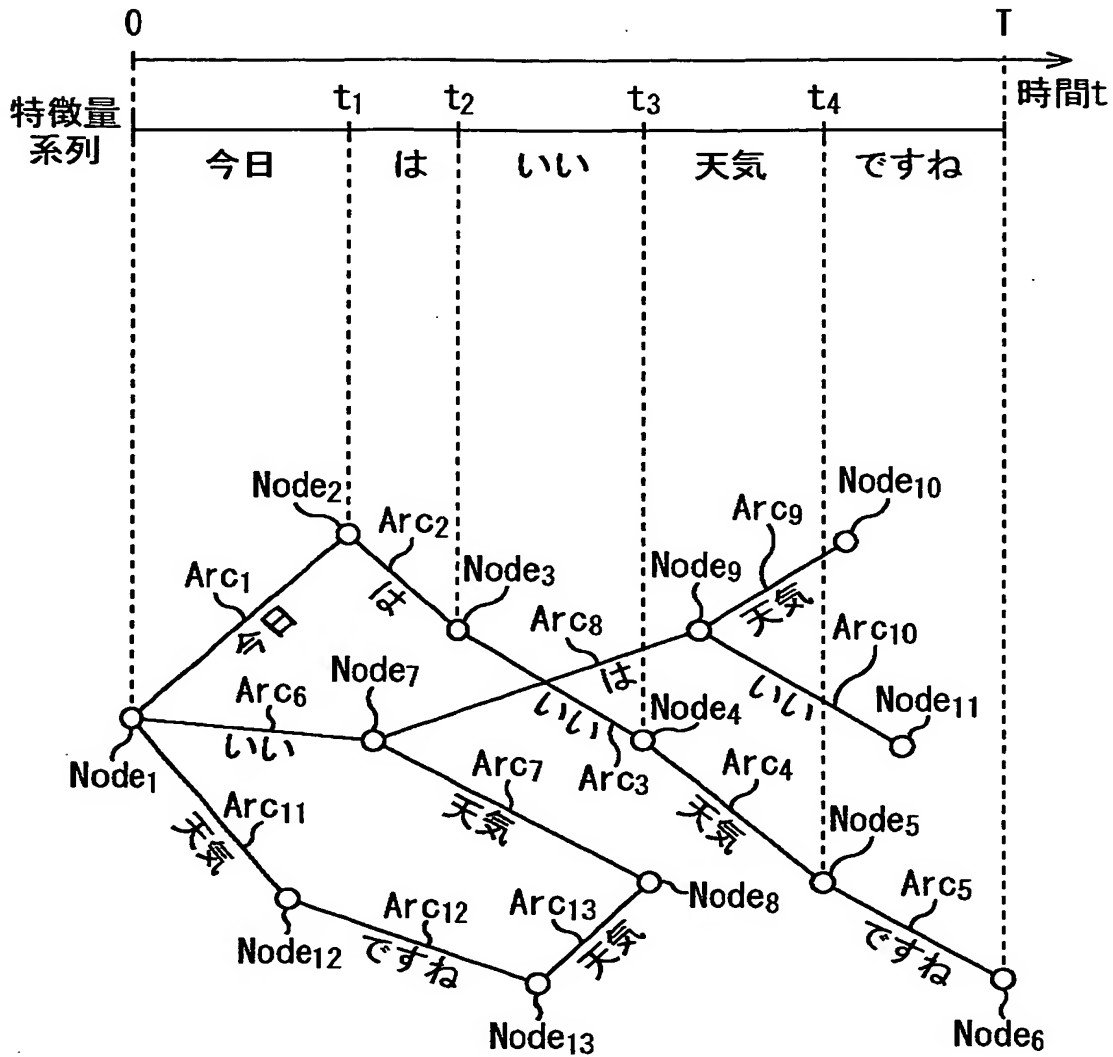


音声認識装置

【図2】

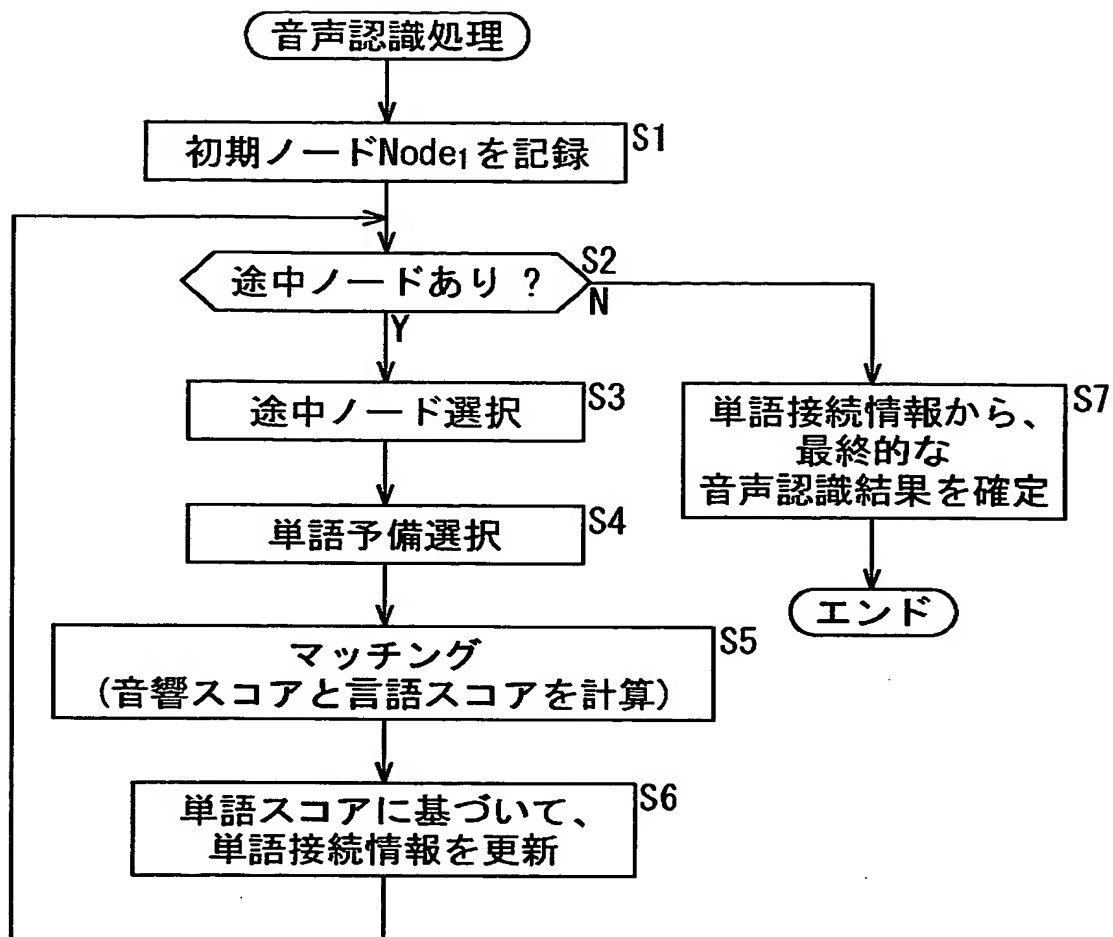


【図3】

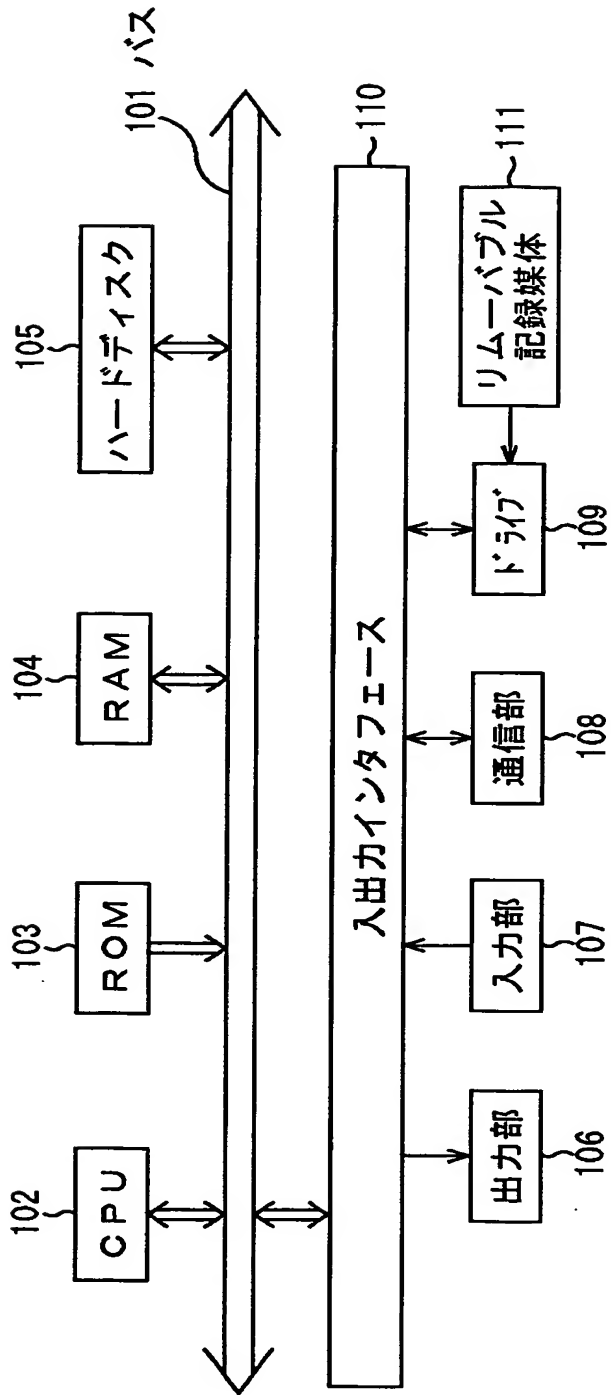


単語接続情報

【図4】



【図5】



コンピュータ

【書類名】 要約書

【要約】

【課題】 リソースの増加を抑えて、音声認識精度を向上させる。

【解決手段】 単語予備選択部 1 3 において、音響スコアおよび言語スコアに基づいて、音声認識結果として確からしい単語が選択される他、例えば、日本語の助詞や助動詞、英語の冠詞や前置詞、その他の音韻数が短い、音響的な特徴量が不安定な単語も選択される。そして、マッチング部 1 4 において、それらの単語を対象に、マッチング処理が行われる。

【選択図】 図 2

出 願 人 履 歴 情 報

識別番号 [000002185]

1. 変更年月日 1990年 8月30日

[変更理由] 新規登録

住 所 東京都品川区北品川6丁目7番35号

氏 名 ソニー株式会社

THIS PAGE BLANK (USPTO)